GLM for Multinomial Outcomes

Multinomial outcomes are much akin to binomial outcomes, with added complexity due to outcomes with more than 2 levels. In such cases it can be difficult to determine an 'order' to the outcomes.

Log-linear models can be used for analysis of this type of data. For example, if we have some distribution where π_{ij} represents the probability that the i^{th} individual falls under the j^{th} category. Assuming the response categories are mutually exclusive:

$$\sum_{j=1}^{J} \pi_{i,j} = 1$$

For each individual i; That is the probabilities add up to 1 for each individual, so we only have J - 1 probability parameters.

Multinomial Distribution

This is an extension of the binomial distribution involving joint distributions. Specifically each trial can result in any of the k events (E1, E2 ... Ek), each with its own respective probabilities. The probability that in n trials we observe Y1 of the E1 outcomes, y2 of the R2 outcomes ... and yk of the Rk outcomes in n independent trials is:

$$P\left(Y_1 = y_1, Y_2 = y_2, \cdots, Y_k = y_k | \pi_1, \pi_2, \cdots, \pi_k\right) = \binom{n}{y_1, y_2, \cdots, y_k} \pi_1^{y_1} \cdot \pi_2^{y_2} \cdots \pi_k^{y_k}$$

with y1 + y1 .. + yk = n and the probabilities summing to 1

The multinomial term:

$$\binom{n}{y_1, y_2, \cdots, y_k} = \frac{n!}{y_1! \cdot y_2! \cdots y_k!}$$

represents the number of possible divisions of n distinct objects into k distinct groups of sizes y1, y2... yk

Consider an example with k = 3 (three possible outcomes):

Mean:

$$E\left(Y_1, Y_2,\ Y_3 | \pi_1, \pi_2, \pi_3\right) = (n\pi_1, n\pi_2, n\pi_3)$$

Variance:

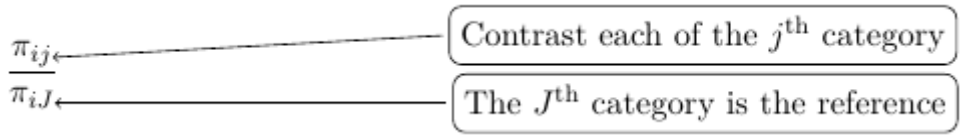
$$\operatorname{Var}\left(Y_1, Y_2,\ Y_3 | \pi_1, \pi_2, \pi_3\right) = \begin{pmatrix} n\pi_1\left(1-\pi_1\right) & -n\pi_1\pi_2 & -n\pi_1\pi_3 \\ -n\pi_1\pi_2 & n\pi_2\left(1-\pi_2\right) & -n\pi_3\pi_2 \\ -n\pi_1\pi_3 & -n\pi_3\pi_2 & n\pi_3\left(1-\pi_3\right) \end{pmatrix}$$

> The multinomial distribution model counts are negatively correlated; The dispersion parameter is $\varphi = 1$

The binomial distribution can be thought of as a particular case of the multinomial distribution with k = 2. We want models for the mean of the response or equivalently for the probabilities, where the probabilities depend on a vector of covariates.

Link Function - Generalized Logits

Perhaps the simplest approach to the Multinomial regression is to designate one of the response categories as the *reference* category and calculate the log-odds for all other categories relative to the reference. For the binomial example, the reference cell is often the 'non-event' category.



So for a multinomial with J categories there are J - 1 distinct odds. For example, in a case with 3 response categories we could use the last category as the reference and hence get 2 generalized odds also called **generalized logits**:

$$\log\left(\frac{\pi_{i1}}{\pi_{i3}}\right) = \beta_0 + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k$$

$$\log\left(\frac{\pi_{i2}}{\pi_{i3}}\right) = \gamma_0 + \gamma_1 X_i^1 + \gamma_2 X_i^2 + \ldots + \gamma_k X_i^k$$

This is kind of like running 2 separate logistic regressions and estimating two separate sets of parameters, but it is a single model and all parameters are estimated by the MLE.

We can solve for the probabilities as a function of the slope parameters (beta and gamma):

$$\pi_{i1} = \frac{\exp\left\{\beta_0 + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k\right\}}{1 + \exp\left\{\beta_0 + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k\right\} + \exp\left\{\gamma_0 + \gamma_1 X_i^1 + \gamma_2 X_i^2 + \ldots + \gamma_k X_i^k\right\}}$$

$$\pi_{i2} = \frac{\exp\left\{\gamma_0 + \gamma_1 X_i^1 + \gamma_2 X_i^2 + \ldots + \gamma_k X_i^k\right\}}{1 + \exp\left\{\beta_0 + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k\right\} + \exp\left\{\gamma_0 + \gamma_1 X_i^1 + \gamma_2 X_i^2 + \ldots + \gamma_k X_i^k\right\}}$$

$$\pi_{i3} = \frac{1}{1 + \exp\left\{\beta_0 + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k\right\} + \exp\left\{\gamma_0 + \gamma_1 X_i^1 + \gamma_2 X_i^2 + \ldots + \gamma_k X_i^k\right\}}$$

Thus, for a multinomial response with J categories, the J - 1 generalized logits are:

$$\begin{aligned}
\log\left(\frac{\pi_{i1}}{\pi_{iJ}}\right) &= \beta_0^1 + \beta_1^1 X_i^1 + \beta_2^1 X_i^2 + \ldots + \beta_k^1 X_i^k \\
\log\left(\frac{\pi_{i2}}{\pi_{iJ}}\right) &= \beta_0^2 + \beta_1^2 X_i^1 + \beta_2^2 X_i^2 + \ldots + \beta_k^2 X_i^k \\
&\vdots \quad \vdots \quad \vdots \\
\log\left(\frac{\pi_{i(J-1)}}{\pi_{iJ}}\right) &= \beta_0^{J-1} + \beta_1^{J-1} X_i^1 + \beta_2^{J-1} X_i^2 + \ldots + \beta_k^{J-1} X_i^k
\end{aligned}$$

And then maximum likelihood can be used to estimate all parameters. For the intercept and each

predictor, we will be estimating J - 1 parameters. Note that it really makes no difference which category is the reference, as we can always convert from one model formulation to another.

Interpretation

The generalized odds of having a response j relative to J are:

$$\exp\left\{\beta_1^j\left(X_i^1 - X_{i'}^1\right) + \beta_2^j\left(X_i^2 - X_{i'}^2\right) + \ldots + \beta_k^j\left(X_i^k - X_{i'}^k\right)\right\}$$

times higher for an individual with characteristics Xi1, Xi2, Xi3... Xik than for a individual with characteristics Xi'1, Xi'2, Xi'3... Xi'k. In particular, keeping X2 through Xk constant for each increase of one unit in X1 the odds of having a response j relative to J changes by a multiplicative factor of exp(beta1j)

Multinomial Ordinal Responses

Consider a multinomial random varaible Yi that may take one of J outcomes, assume there is a natural order from 1 to J. The probability falls into category j or lower is f_{ij} = P(Y_i <= j), also called the *cumulative probabilities.*

When we have a clear order between categories several alternative regression models are available. We could use generalized logits model but it ignores the order of the levels and this have diminished power. The drawback of multinomial logits compared to general multinomial logits is that assumptions imposed may not be valid. We consider 3 types of models in which the order of the factors is relevant:

1. Cumulative Logits models - most widely used

$$L_{ij} = \log\left(\frac{f_{ij}}{1 - f_{ij}}\right) = \log\left(\frac{\pi_{i,1} + \pi_{i,2} + \pi_{i,3} + \ldots + \pi_{i,j}}{\pi_{i,j+1} + \pi_{i,j+2} + \pi_{i,j+3} + \ldots + \pi_{i,J}}\right), \text{ for } j = 1, 2, 3, \ldots, J-1$$

2. Adjacent-Categories Logits models

$$L_{ij} = \log\left(\frac{\pi_{i,j}}{\pi_{i,j+1}}\right), \text{ for } j = 1, 2, 3, \ldots, J-1$$

3. Continuation-Ratio Logits models

$$L_{ij} = \log\left(\frac{\pi_{i,j}}{\pi_{i,j+1} + \pi_{i,j+2} + \pi_{i,j+3} + \ldots + \pi_{i,J}}\right), \text{ for } j = 1,\ 2,\ 3,\ \ldots, J-1$$

Once one decides what type of logit to use, we construct the logits model with the following identities:

$$L_{ij} = \alpha_j + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k \text{ for } j = 1,\ 2,\ 3,\ \ldots,\ J-1$$

Where L_{ij} are one of the 3 logit types considered above.

These models assume separate intercept for each logit j and the same coefficients for the covariates. The intercept parameters (alpha) are also called cut-point parameters and follow the order of the ordered categorical outcome.

Type of Logits	*Logits* (for a multinomial outcome with 4 categories)
Cumulative	$\log\left(\frac{\pi_1}{\pi_2+\pi_3+\pi_4}\right), \log\left(\frac{\pi_1+\pi_2}{\pi_3+\pi_4}\right), \log\left(\frac{\pi_1+\pi_2+\pi_3}{\pi_4}\right)$
Adjacent	$\log\left(\frac{\pi_1}{\pi_2}\right), \log\left(\frac{\pi_2}{\pi_3}\right), \log\left(\frac{\pi_3}{\pi_4}\right)$
Continuation-ratio	$\log\left(\frac{\pi_1}{\pi_2+\pi_3+\pi_4}\right), \log\left(\frac{\pi_2}{\pi_3+\pi_4}\right), \log\left(\frac{\pi_3}{\pi_4}\right)$

Cumulative Logits Models

A model for the cumulative logits L_{ij} is a model for a binary outcome in which success is defined as **any category <= j.** Note that the cumulative logits model for a multinomial outcome is not equivalent with J-1 classical logit models. We can also write it as:

$$L_{ij} = \log\left(\frac{P\left(Y_i \leq j\right)}{1 - P\left(Y_i \leq j\right)}\right) = \log\left(\frac{f_{ij}}{1 - f_{ij}}\right)$$

The model also assumes for 2 individuals the relative odds of being at or below a given cut-point (j) on the outcome is independent of the choice of cut-point (j); Meaning separate intercepts for each j of the ordered categorical outcome and the same slope for each predictor. As a consequence of this two individuals i and i' are independent of j:

$$L_{ij} - L_{i'j} = \log\left(\frac{\frac{P(Y_i \leq j)}{1-P(Y_i \leq j)}}{\frac{P\left(Y_{i'} \leq j\right)}{1-P\left(Y_{i'} \leq j\right)}}\right) = \log\left(\text{OR}\{i \text{ vs } i^{'}\}\right) = \beta_1\left(X_i^1 - X_{i'}^1\right) + \beta_2\left(X_i^2 - X_{i'}^2\right) + \ldots + \beta_k\left(X_i^k - X_{i'}^k\right)$$

Because of this, this model is also called the **proportional odds model.**

The proportional cumulative odds model will estimate a common value for two odds ratios. While the calculations might not always be exactly equal, the estimated values should suggest proportional odds. This can be tested in a logistic model in SAS.

Interpretation

The intercept alpha represents the log odds of having a response <= j when all explanatory variables are 0.

The slope is the increase in logodds of having a response <= j for an individual with characteristics X1 ... Xk than for an individual with characteristics X'1 ... X'k. Particularly it means keeping X2 ... Xk constant each increase of X1 by one unit increases the log odds by beta.

The odds of having a response <= j is:

$$\exp\left\{\beta_1\left(X_i^1 - X_{i'}^1\right) + \beta_2\left(X_i^2 - X_{i'}^2\right) + \ldots + \beta_k\left(X_i^k - X_{i'}^k\right)\right\}$$

Each explanatory variable specific odds ratio can be interpreted as the effect of that variable on the ODDS of being in a lower rather than higher category.

Adjacent Categories Logits Models

$$L_{ij} = \log\left(\frac{\pi_{ij}}{\pi_{i(j+1)}}\right) = \alpha_j + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k \text{ for } j = 1,\ 2,\ 3,\ \ldots,\ J-1$$

The same linear combination of predictors applies simultaneously for all J-1 pairs of adjacent response categories. The only thing that varies is the intercept.

Sometimes adjacent logits are defined equivalently as:

$$\log\left(\frac{\pi_{i(j+1)}}{\pi_{ij}}\right) = -\log\left(\frac{\pi_{ij}}{\pi_{i(j+1)}}\right)$$

In SAS PROC CATMOD defines the adjacent logits as the former, log(πi(j+1) / πij)

Parameter Interpretation

For two individuals the relative odds of being in category **j instead of j+1** is independent of choice of cut-point. Thus for 2 individuals i and i':

$$L_{ij} - L_{i'j} = \log\left(\frac{\frac{P(Y_i=j)}{P(Y_i=j+1)}}{\frac{P(Y_{i'}=j)}{P(Y_{i'}=j+1)}}\right) = \log\left(\text{OR}\left(i \text{ vs } i'\right)\right)$$

$$= \alpha_j + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k - (\alpha_j + \beta_1 X_{i'}^1 + \beta_2 X_{i'}^2 + \ldots + \beta_k X_{i'}^k)$$

$$= \beta_1\left(X_i^1 - X_{i'}^1\right) + \beta_2\left(X_i^2 - X_{i'}^2\right) + \ldots + \beta_k\left(X_i^k - X_{i'}^k\right)$$

Again, independent of j. That is the ratio of odds of a response j instead of j+1 comparing two groups are the same for every j.

The adjacent odds are $\exp(\{\beta_1(X_i^1 - X_{i'}^1) + \beta_2(X_i^2 - X_{i'}^2) + \ldots + \beta_k(X_i^k - X_{i'}^k)\})$ times higher for an individual with characteristics X1 ... Xk than for an individual with characteristics X'1 ... X'k. Particularly it means keeping X2-Xk constant each increase of X1 by one unit increases the adjacent odds by exp(beta).

If we allow the slopes to vary, the proportional odds assumption does not hold.

Continuation Ratio Logits Model

$$\log\left(\frac{\pi_{ij}}{\pi_{i(j+1)} + \pi_{i(j+2)} + \ldots + \pi_{iJ}}\right) = \alpha_j + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k \text{ for } j = 1,\ 2,\ 3,\ \ldots,\ J-1$$

The same linear logit effect applies simultaneously for all J-1 continuation-ratio logits, the only coefficients that vary are the intercepts.

Let $\theta_{ij} = P(Y_i = j \mid Y_i \geq j)$ be the probability of the response in the jth category given that we know the response is >= j.

$$\theta_{ij} = P\left(Y_i = j \mid Y_i \geq j\right) = \frac{\pi_{ij}}{\pi_{ij} + \pi_{i(j+1)} + \pi_{i(j+2)} + \ldots + \pi_{iJ}}$$

$$\log\left(\frac{\theta_{ij}}{1-\theta_{ij}}\right) = \left(\frac{\frac{\pi_{ij}}{\pi_{ij}+\pi_{i(j+1)}+\pi_{i(j+2)}+\ldots+\pi_{iJ}}}{\frac{\pi_{ij}+\pi_{i(j+1)}+\pi_{i(j+2)}+\ldots+\pi_{iJ}-\pi_{ij}}{\pi_{ij}+\pi_{i(j+1)}+\pi_{i(j+2)}+\ldots+\pi_{iJ}}}\right) = \log\left(\frac{\pi_{ij}}{\pi_{i(j+1)} + \pi_{i(j+2)} + \ldots + \pi_{iJ}}\right)$$

In other words the continuation-Ratio logits are the ordinary logits of the conditional probabilities of the response being j given that the response is at least j.

Alternatively, some books define the continuation-Ratio logits as:

$$\log\left(\frac{\pi_{i(j+1)}}{\pi_{i1} + \pi_{i2} + \ldots + \pi_{ij}}\right) = \alpha_j + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k \text{ for } j = 1,\ 2,\ 3,\ \ldots,\ J-1$$

This is not included in SAS, needs to be programmed.

This model might be more appropriate when ordered categories represent a progression of stages.

Parameter Interpretation

Based on the model for two individuals, the relative odds of being classified in the category j given it is classified in a *a* >= j category is independent of the choice of cut-point. For two individuals i and i':

$$\begin{aligned} L_{ij} - L_{i'j} = \log\left(\frac{\frac{\theta_{ij}}{1-\theta_{ij}}}{\frac{\theta_{ij'}}{1-\theta_{ij'}}}\right) &= \log\left(\text{OR}\left(i \text{ vs } i'\right)\right) \\ &= \alpha_j + \beta_1 X_i^1 + \beta_2 X_i^2 + \ldots + \beta_k X_i^k - \left(\alpha_j + \beta_1 X_{i'}^1 + \beta_2 X_{i'}^2 + \ldots + \beta_k X_{i'}^k\right) \\ &= \beta_1\left(X_i^1 - X_{i'}^1\right) + \beta_2\left(X_i^2 - X_{i'}^2\right) + \ldots + \beta_k\left(X_i^k - X_{i'}^k\right) \end{aligned}$$

That is the odds of being classified in the jth category given it is classified in a category >= jth comparing two groups is the same for every j.

The continuation ratio odds are $\exp\left(\left\{\beta_1\left(X_i^1 - X_{i'}^1\right) + \beta_2\left(X_i^2 - X_{i'}^2\right) + \ldots + \beta_k\left(X_i^k - X_{i'}^k\right)\right\}\right)$ times higher for an individual with characteristics X1 ... Xk than for an individual with characteristics X'1 ... X'k. Particularly it means keeping X2-Xk constant each increase of X1 by one unit increases the continuation odds by exp(beta).

SAS Code

SAS is a series of crappy procedures with no coherent structure or consistency.

```
title2 'Using PROC LOGISTIC  - continuous age linear and quadratic effects';
proc logistic data=contra order=data ;
freq count;
model method=cage agesq/link=glogit ;
output out=aa(where=(_level_ in (1,2) and method in (1,2) and method=_level_)) xbeta=pred;
run;
```

```
title2 'Using PROC CATMOD - continuous age linear and quadratic effects';
proc catmod data=contra order=data;
direct cage agesq;
weight count;
model method=cage agesq;
run;

/*Using Genmod*/
title2 'Using PROC GENMOD - continuous age linear and quadratic effects';
proc genmod data=contra;
 class method age;
 model count = method age method*cage method*agesq/dist=p;
run;

axis1 label=('AGE') minor=none order=(1 to 7 by 1) value=(tick=1 '15-19' tick=2 '20-24' tick=3 '25-29' tick=4
'30-34' tick=5 '35-39'
tick=6 '40-44' tick=7 '45-49') ;
axis2 label=(a=90 'Log Generalized Logit (None as reference)') minor=none ;
symbol1 v=circle i=j cv=red ci=blue;
symbol2 v=dot i=j cv=green ci=magenta;
legend1 value=('Observed' 'Predicted') label=('') ;
title1 h=1.2 'Predicted vs. Observed For Method=Sterilization';
proc gplot data=aa;
plot elsn*age pred*age/vaxis=axis2 haxis=axis1 overlay  legend=legend1;;
where method=1;
run;

/* Using SGplot */
ods listing gpath="Z:/Documents/BS853/Spring 2019/Class 6/";
ods graphics / reset width=600px height=400px  imagename="Pred1" imagefmt=gif;
title1 h=1.2 'Predicted vs. Observed For Method=Sterilization';
proc sgplot data=aa cycleattrs;
 yaxis grid label="Log Generalized Logit (None as reference)";
  xaxis grid label="Age";
   series  x=age y=elsn / lineattrs = (pattern=solid thickness=3)
          legendlabel = "Observed Data" name= "elsn";
   scatter x=age y=elsn/ markerattrs=(symbol=circlefilled size=10  color=orange);;
   series  x=age y=pred / lineattrs = (pattern=solid thickness=3)
          legendlabel = "Model" name= "pred";
```

```
   scatter x=age y=pred/ markerattrs=(symbol=squarefilled size=10 color=green);
 keylegend "elsn" "pred";
where method=1;
run;

title1 h=1.2 'Predicted vs. Observed For Method=Other';
proc gplot data=aa;
plot elsn*age pred*age/vaxis=axis2 haxis=axis1 overlay  legend=legend1;;
where method=2;
run;

/* Using SGplot */
ods listing gpath="Z:/";
ods graphics / reset width=600px height=400px  imagename="Pred2" imagefmt=gif;
title1 h=1.2 'Predicted vs. Observed For Method=Other';
proc sgplot data=aa cycleattrs;
 yaxis grid label="Log Generalized Logit (None as reference)";
  xaxis grid label="Age";
   series  x=age y=elsn / lineattrs = (pattern=solid thickness=3)
          legendlabel = "Observed Data" name= "elsn";
   scatter x=age y=elsn/ markerattrs=(symbol=circlefilled size=10  color=orange);;
   series  x=age y=pred / lineattrs = (pattern=solid thickness=3)
          legendlabel = "Model" name= "pred";
   scatter x=age y=pred/ markerattrs=(symbol=squarefilled size=10 color=green);
 keylegend "elsn" "pred";
where method=2;
run;



/*Schoolchildren learning style preference and School program*/
/* styles 1=Self,2=Team,3=Class*/
data school;
 do style=1 to 3;
input school program $ count @@;
output;
end;
cards;
1 Regular 10 1 Regular  17 1 Regular  26
1 After    5 1 After    12 1 After    50
2 Regular 21 2 Regular  17 2 Regular  26
```

```
2 After   16 2 After    12 2 After    36
3 Regular 15 3 Regular  15 3 Regular  16
3 After   12 3 After    12 3 After    20
run;

title1 'Estimating parameters using contrast statement';
options ls=97 nodate pageno=1;
ods select contrastestimate;
proc logistic data=school order=data;
freq count;
class school program/param=glm;
model style=school program /link=glogit;
contrast 'reg vs. after' program -1 1/estimate=exp;
run;

/*Consider data on association of cumulative incidence of
pneumoconiosis and years working at a coalmine.*/

data ph;
do type=1 to 3;
input age count @@;
output;
end;
cards;
1 218 1 13 1 12
2  71 2 25 2 32
run;

options ls=97 nodate pageno=1;
title1 'Cumulative Logit models - PROC GENMOD';
  proc genmod data=ph order=data descending;
      freq count;  /*<-group data*/
      class age;
      model type = age / dist=multinomial link=cumlogit type3;
      estimate 'LogOR12' age -1 1 / exp;
   run;
title1 'Cumulative Logit models - PROC LOGISTIC';
     proc logistic data=ph order=data descending;
      freq count;
      class age/param=glm;
```

```
      model type = age;
     contrast 'LogOR12' age -1 1 / estimate=exp;
   run;

/* Metal Impairement data - Agresti, 1990*/

proc import datafile="Z:\Documents\BS853\Spring 2015\Class 6\impair.xls"
out = mimpair replace;
run;

options ls=97 nodate pageno=1;
title1 'Cumulative Logit models - PROC GENMOD';
  proc genmod data=mimpair rorder=data;
      model level = SES EVENTS/ dist=multinomial link=cumlogit type3;
      estimate 'LogOR12' SES 1  / exp;
   run;

title1 'Cumulative Logit models - PROC LOGISTIC';
    proc logistic data=mimpair order=data ;
     class level;
      model level = SES EVENTS ;
      contrast 'LogOR12' SES 1  / estimate=exp;
   run;
```

The RESPONSE statement in PROC CATMOD defines the function of the response probabilities (the link function) that we want to model as linear combinations of the predictors. Operations in the RESPONSE statement should be read right to left with the vector of observed response probabilities understood to be the right-most quantity. The rows of the matrices are separated by commas.

The _RESPONSE_ variable is a categorical variable with seperate levels for each of the logits.

```
/*satisfaction with THR and hospital volume and year of surgery.*/
proc format ;
  value vol 1='High' 0='Low';
run;
data thr;
input  year highvol satisf count;
format highvol vol.;
datalines;
1       0       1       84
1       0       2      231
1       0       3       99
```

```
1       1       1      473
1       1       2      493
1       1       3       93
2       0       1      150
2       0       2      347
2       0       3      117
2       1       1      332
2       1       2      387
2       1       3       62
3       0       1      257
3       0       2      889
3       0       3      234
3       1       1      571
3       1       2      793
3       1       3      112
;
run;

options pageno=1 nodate ps=57 ls=91;
title1 'Adjacent Logits models for THR';
proc catmod data=thr;
  weight count;
  response ALOGIT;
  model satisf= _response_ highvol year;
run;
quit;



/* let the effect of year vary with the cutpoint*/
title1 'Adjacent Logits models for THR - with coefficents dependent on the level of satisfcation';
proc catmod data=thr ;
  weight count;
  response ALOGIT;
  model satisf= _response_ highvol _response_|year;
run;
quit;



/* number of hip fractures by race age and BMI */
/* n=0 then no hip fractures, n=1 then 1 hip fractures, n=2 then more than 1 hip fractures */
data nhf;
```

```
do n=0 to 2;
input race $ age $ bmi count @@;
output;
end;
cards;
white□<=80 □1□39□white□<=80 □1□29□white□<=80 □1□8
white□<=80□2□4□white□<=80□2□8□white□<=80□2□1
white□<=80□3□11□white□<=80□3□9□white□<=80□3□6
white□<=80□4□48□white□<=80□4□17□white□<=80□4□8
white□>80 □1□231□white□>80 □1□115□white□>80 □1□51
white□>80□   2□17□white□>80□   2□21□white□>80□   2□13
white□>80□   3□18□white□>80□   3□28□white□>80□   3□45
white□>80□   4□197□white□>80□   4□111□white□>80□   4□35
black□<=80 □1□19□black□<=80 □1□40□black□<=80 □1□19
black□<=80□2□5□black□<=80□2□17□black□<=80□2□7
black□<=80□3□2□black□<=80□3□14□black□<=80□3□3
black□<=80□4□49□black□<=80□4□79□black□<=80□4□24
black□>80 □1□110□black□>80 □1□133□black□>80 □1□103
black□>80□   2□18□black□>80□   2□38□black□>80□   2□25
black□>80□   3□11□black□>80□   3□25□black□>80□   3□18
black□>80□   4□178□black□>80□   4□206□black□>80□   4□81
;
run;
options pageno=1 nodate ps=57 ls=91;
/* continuation ratio logits models*/
 title1 'Continuation-ratio Logits models for NHF using the response statement';
proc catmod data=nhf;
    weight count;
    response * 1 -1 0  0,
            0  0 1 -1  log  * 1 0 0,
                       0 1 1,
                       0 1 0,
                       0 0 1;  /* p1
                                  p2
                                  p3  */
    model n=_response_ age bmi|race/  predict;
    run;
    quit;

□ /* Use second way to defining continuation ration logits */
title1 'Continuation-ratio Logits models for NHF using the response statement(2)';
```

```
  proc catmod data=nhf;
     weight count;
     response * -1 1  0 0,
               0 0 -1 1  log  * 1 0 0,
                           0 1 0,
                           1 1 0,
                           0 0 1;
     model n=_response_ age bmi|race/ prob predict;
     run;
     quit;

/* let the coefficients depend on the number of hip fracturs */
 title2 'Coefficients dependent on the number of HF';
proc catmod data=nhf;
     weight count;
     response * 1 -1 0 0,
              0 0 1 -1  log  * 1 0 0,
                           0 1 1,
                           0 1 0,
                           0 0 1;

     model n=_response_|bmi _response_|age _response_|race  bmi|race/ freq;
     run;
     quit;

title1 'Adjacent Logits models for THR using the response statement';
proc catmod data=thr;
  weight count;
 response * -1  1 0,
            0 -1 1  log  * 1 0 0,
                        0 1 0,
                        0 0 1;  /* p1
                                   p2
                                   p3  */
  model satisf= _response_ highvol year;
run;
quit;

title1 'cumulative Logits models for THR using the response statement';
proc catmod data=thr;
  weight count;
```

```
 response *  1 -1 0  0,
          0  0 1 -1  log  * 1 0 0,
                         0 1 1,
                         1 1 0,
                         0 0 1;  /* p1
                                 p2
                                 p3  */
  model satisf= _response_ highvol year;
run;
quit;
title1 'cumulative Logits models for THR using Proc Logistic ';
proc logistic data=THR;
  class highvol year;
  model satisf= highvol year;
  freq count;
run;

/*****************************************************************************/
/***              Overdispersion in binomial models                        ***/
/*****************************************************************************/
DATA aggregate;
  INPUT pot total yes cultivar soil @@;
cards;
 1 16  8 0 0   2 51 26 0 0
 5 36  9 0 0   6 81 23 1 0
 9 28  8 1 0  10 62 23 1 0
13 41 22 0 1  14 12  3 0 1
17 30 15 1 1  18 51 32 1 1
 3 45 23 0 0   4 39 10 0 0
 7 30 10 1 0   8 39 17 1 0
11 51 32 0 1  12 72 55 0 1
15 13 10 0 1  16 79 46 1 1
19 74 53 1 1  20 56 12 1 1
;
run;



/* no scale - aggregated data'*/
options pageno=1 nodate ps=57 ls=91;
title1 ' No Scale ';
proc genmod data=aggregate;
```

```
model yes/total = cultivar|soil/type3 dist=binomial ;
run;

/*  scale - aggregated data'*/
/* Model 2 */
title1 ' Scale parameter estimated by the scaled dispersion ';
proc genmod data=aggregate;
model yes/total = cultivar|soil/type3 dist=binomial dscale ;
run;
```

Revision #7
Created 14 March 2023 18:18:04 by Elkip
Updated 21 March 2023 20:21:56 by Elkip